Data Plane Development Kit

Quality of Service (QoS)

Cristian Dumitrescu
SW Architect - Intel

Apr 21, 2015

Legal Disclaimer

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm%20 Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit Intel Performance Benchmark Limitations

All products, computer systems, dates and figures specified are preliminary based on current expectations, and are subject to change without notice.

Celeron, Intel, Intel logo, Intel Core, Intel Inside, Intel Inside logo, Intel. Leap ahead., Intel. Leap ahead. logo, Intel NetBurst, Intel SpeedStep, Intel XScale, Itanium, Pentium, Pentium Inside, VTune, Xeon, and Xeon Inside are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Intel® Active Management Technology requires the platform to have an Intel® AMT-enabled chipset, network hardware and software, as well as connection with a power source and a corporate network connection. With regard to notebooks, Intel AMT may not be available or certain capabilities may be limited over a host OS-based VPN or when connecting wirelessly, on battery power, sleeping, hibernating or powered off. For more information, see http://www.intel.com/technology/iamt.

64-bit computing on Intel architecture requires a computer system with a processor, chipset, BIOS, operating system, device drivers and applications enabled for Intel® 64 architecture. Performance will vary depending on your hardware and software configurations. Consult with your system vendor for more information.

No computer system can provide absolute security under all conditions. Intel® Trusted Execution Technology is a security technology under development by Intel and requires for operation a computer system with Intel® Virtualization Technology, an Intel Trusted Execution Technology-enabled processor, chipset, BIOS, Authenticated Code Modules, and an Intel or other compatible measured virtual machine monitor. In addition, Intel Trusted Execution Technology requires the system to contain a TPMv1.2 as defined by the Trusted Computing Group and specific software for some uses. See http://www.intel.com/technology/security/ for more information.

†Hyper-Threading Technology (HT Technology) requires a computer system with an Intel® Pentium® 4 Processor supporting HT Technology and an HT Technology-enabled chipset, BIOS, and operating system. Performance will vary depending on the specific hardware and software you use. See www.intel.com/products/ht/hyperthreading_more.htm for more information including details on which processors support HT Technology.

Intel® Virtualization Technology requires a computer system with an enabled Intel® processor, BIOS, virtual machine monitor (VMM) and, for some uses, certain platform software enabled for it. Functionality, performance or other benefits will vary depending on hardware and software configurations and may require a BIOS update. Software applications may not be compatible with all operating systems. Please check with your application vendor.

* Other names and brands may be claimed as the property of others.

Other vendors are listed by Intel as a convenience to Intel's general customer base, but Intel does not make any representations or warranties whatsoever regarding quality, reliability, functionality, or compatibility of these devices. This list and/or these devices may be subject to change without notice.

Status

- DPDK QoS solution is out there since April '13 (Release 1.4)
- Traffic Management for 40GbE line rate (128-byte or larger pkts) is typically achieved with just 2x Intel Xeon CPU cores

Traffic Metering

- ✓ Single Rate Three Color Marker (srTCM)
- ✓ Two Rate Three Color Marker (trTCM)

Traffic Management

- ✓ Hierarchical scheduler: 5-level hierarchy, 64K (or more) packet queues
- ✓ Traffic shaping, strict priority, byte-level WRR (WFQ)
- ✓ Fairness in case of oversubscription
- ✓ Congestion management: tail drop, RED, WRED

DPDK Map for QoS

- Libraries (lib folder): librte_meter, librte_sched
- Sample applications (examples folder): qos_meter, qos_sched
- Documentation:
  - ✓ DPDK Programmer's Guide, Chapter "Quality of Service (QoS) Framework"
  - ✓ DPDK Sample Applications User Guide, Chapters "QoS Metering Sample Application" and "QoS Scheduler Sample Application"

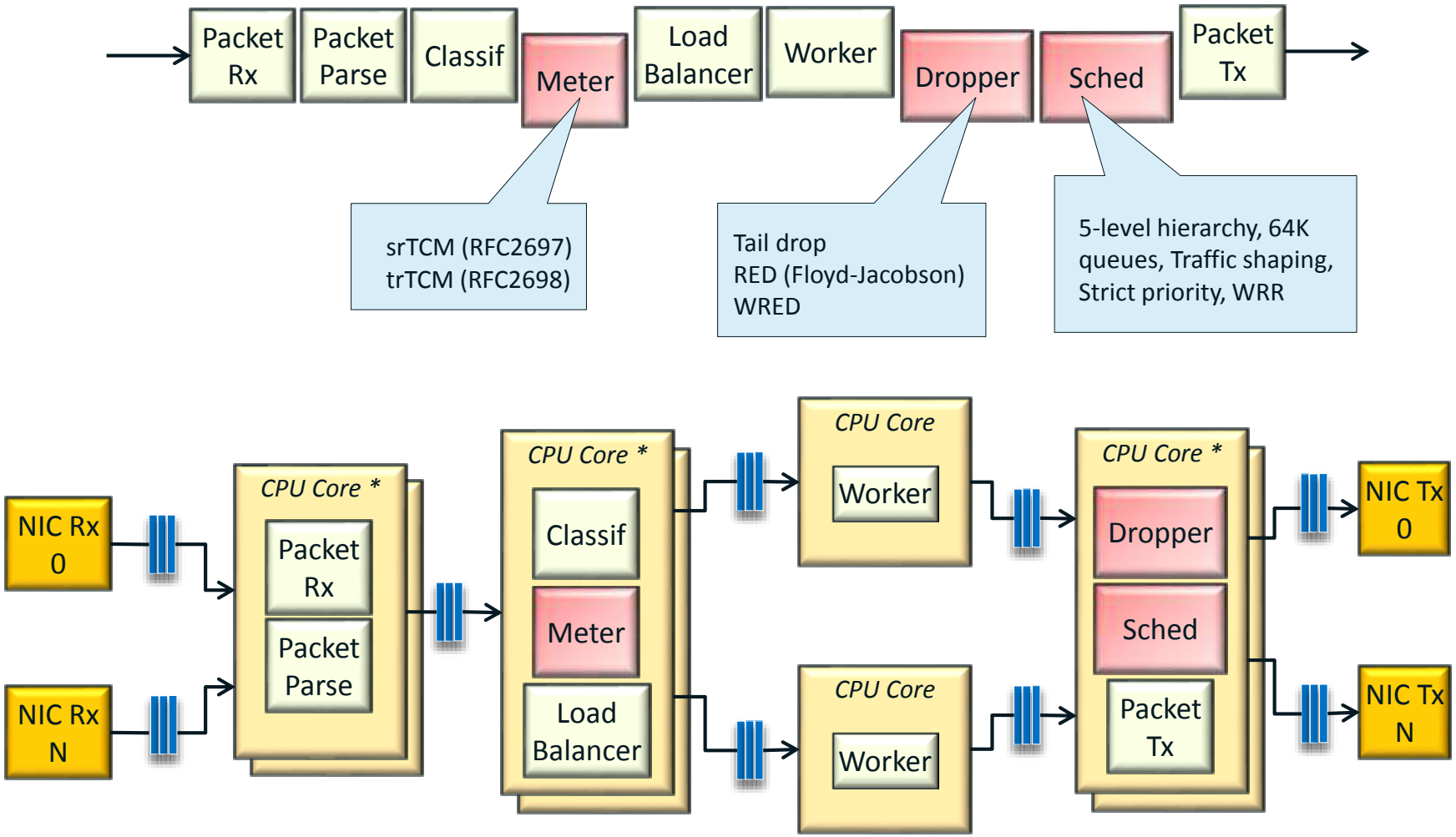
Functional Overview

Why Quality of Service (QoS) ?

Not all packets are created equal …

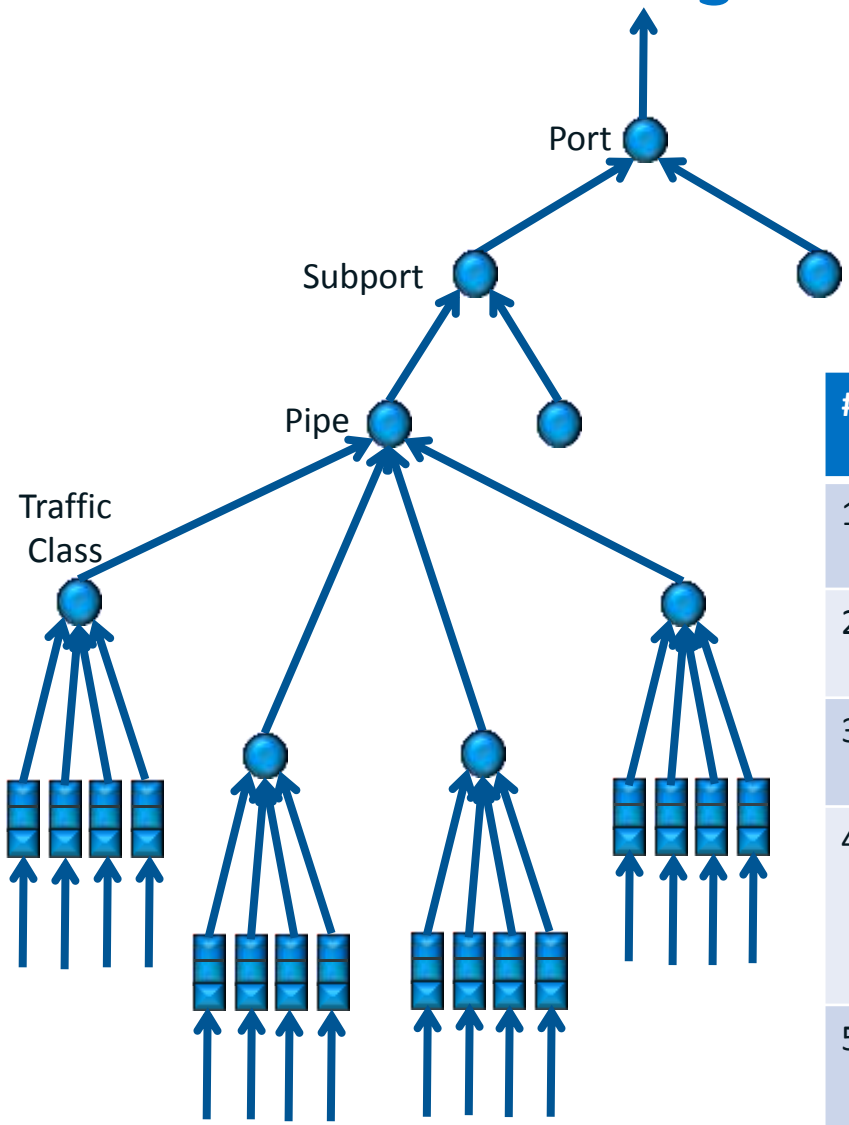
- Different users have different needs
  - Need to enforce per user Service Level Agreement (SLA)
- Different traffic types have different needs
  - Voice, video and data streaming have different requirements in terms of acceptable delay, delay variation (jitter) and packet loss rate

DPDK Packet Pipeline with QoS



* Number of CPU cores depends on feature set and performance target of each application

DPDK Scheduling Hierarchy per Output Port



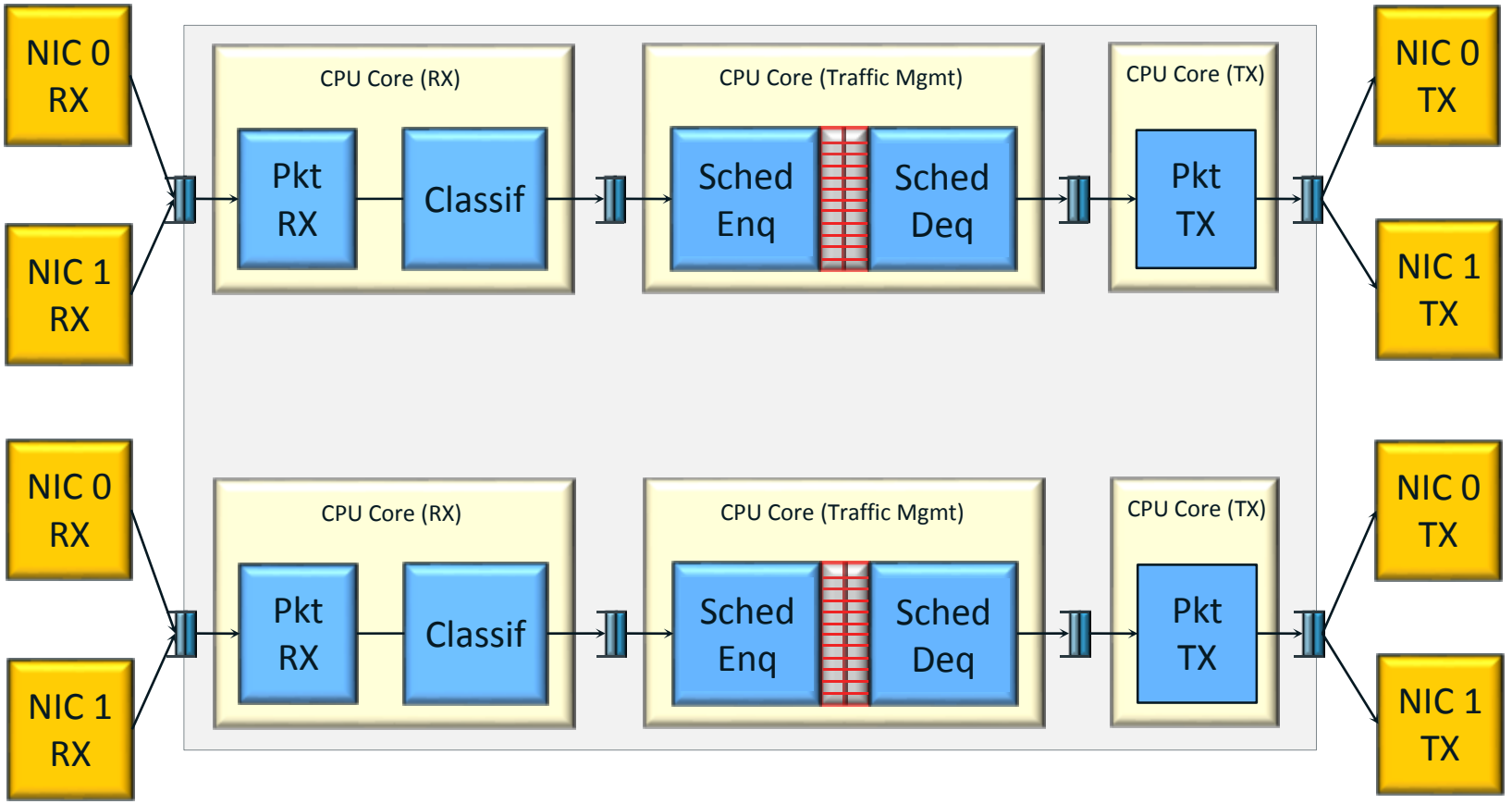
#	Level Name	Siblings per parent	Functional description
1	Port	Config	Output Ethernet port (1GbE/10GbE)
2	Subport	Config (Typ: <8)	Traffic shaping (token bucket per subport)
3	Pipe	Config (Typ: 4K)	Traffic shaping (token bucket per pipe)
4	Traffic Class	4	TCs of the same pipe serviced in strict priority; Upper limit enforced; BW reuse within same pipe
5	Queue	4	Queues of the same TC are serviced using byte-level WRR (i.e. WFQ)

Functional Testing for Accuracy

#	Test Objective	Status
1.	**Traffic shaping for subport**. Check that subport rate limiting is *accurate*.	*Pass*
2.	**Subport TC oversubscription**. On subport TC oversubscription, check that subport member pipe rate allowance is *accurately* enforced.	*Pass*
3.	**Traffic shaping for pipe**. Check that pipe rate limiting is *accurate*.	*Pass*
4.	**Subport Traffic Classes**. Check that subport TC rate limiting is *accurate*.	*Pass*
5.	**Pipe Traffic Classes**. Check that pipe TC rate limiting is *accurate*.	*Pass*
6.	**Strict priority**. Check that pipe TCs are scheduled in strict priority.	*Pass*
7.	**BW reuse within the pipe**. Check that lower priority TCs are able to reuse pipe BW not used by higher priority TCs.	*Pass*
8.	**Byte-level WRR**. Check that pipe TC queue weight-based scheduling is *accurate*.	*Pass*

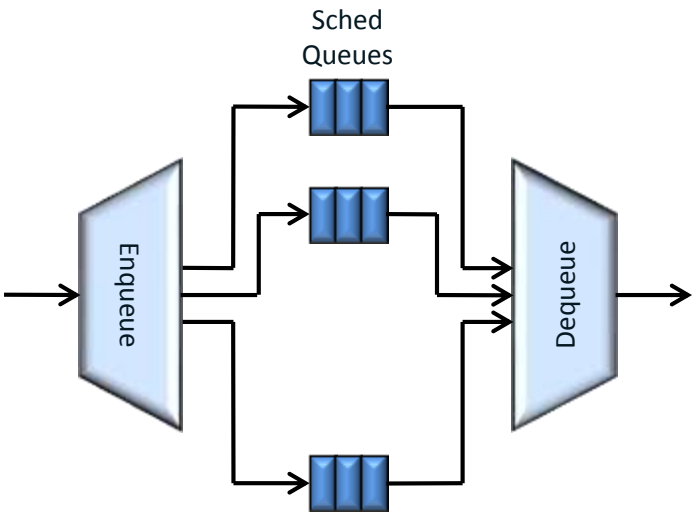
Ensuring functional accuracy for hierarchical scheduler features is top priority.

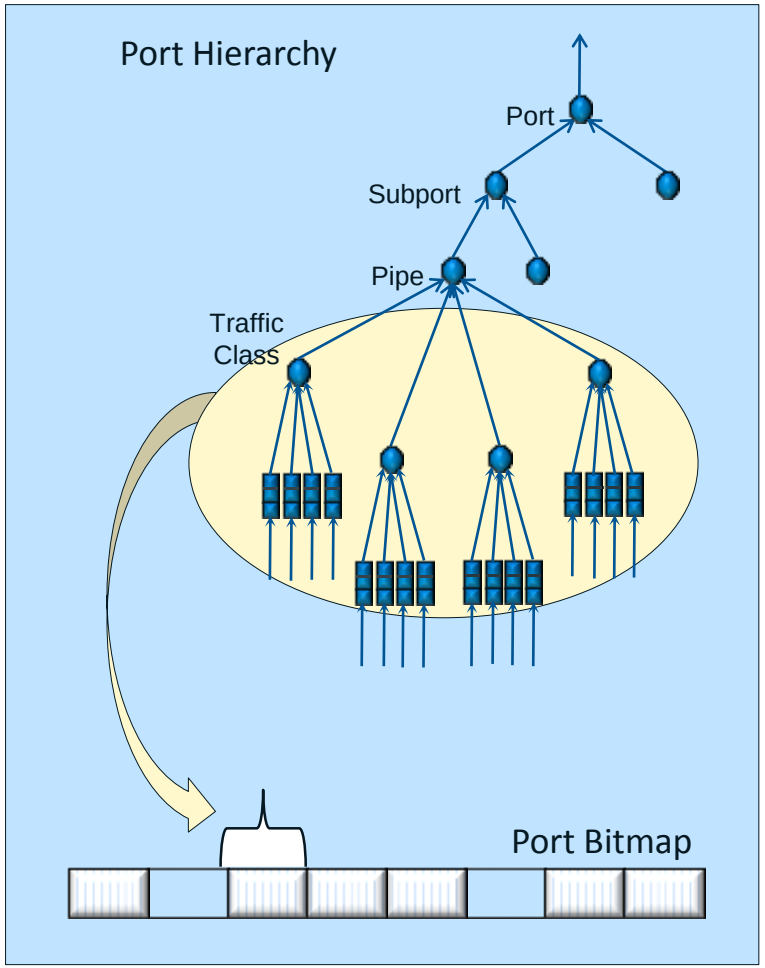
Hierarchical Scheduler Example Application



Implementation Overview

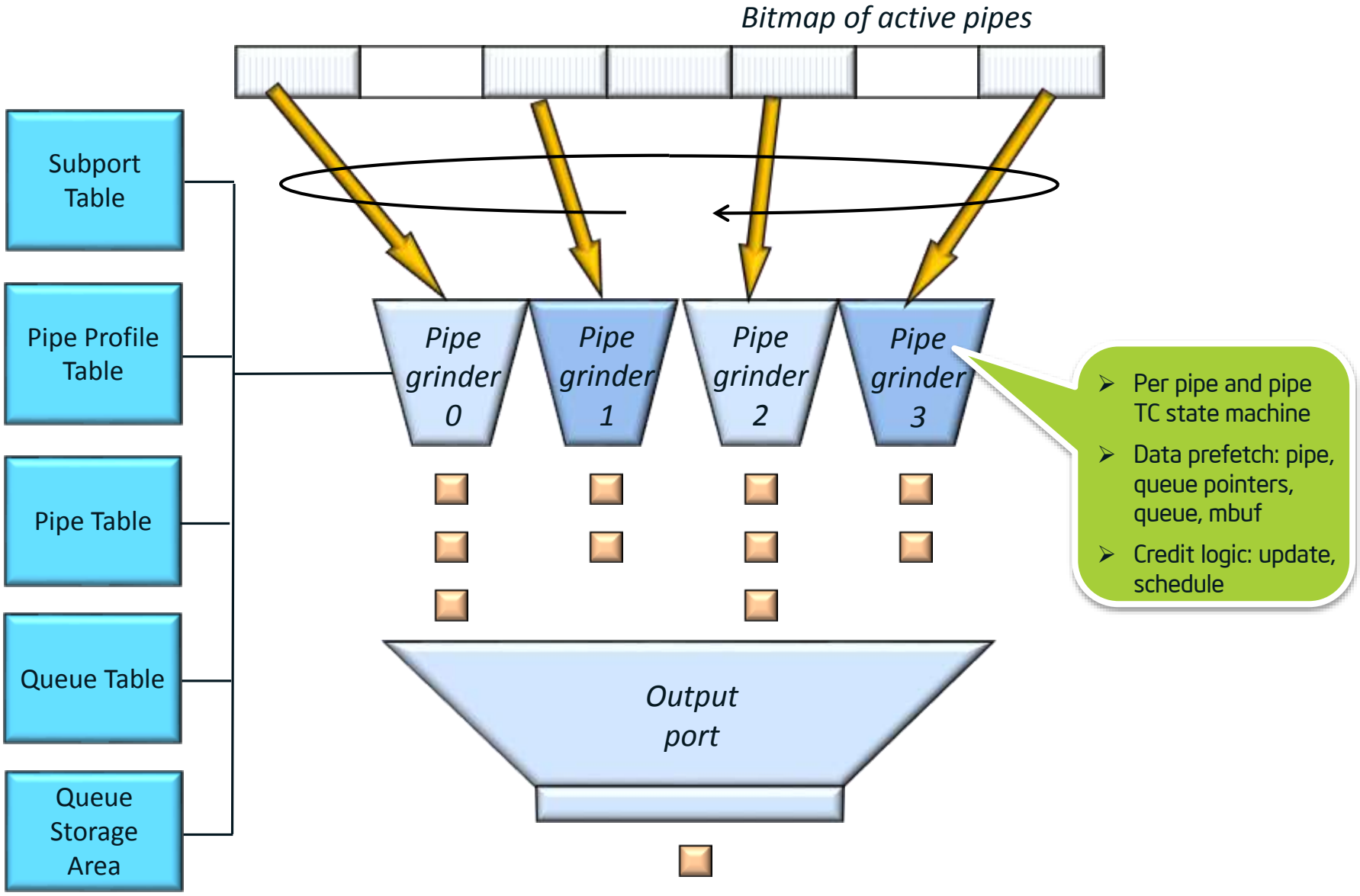
Tracking active queues / pipes



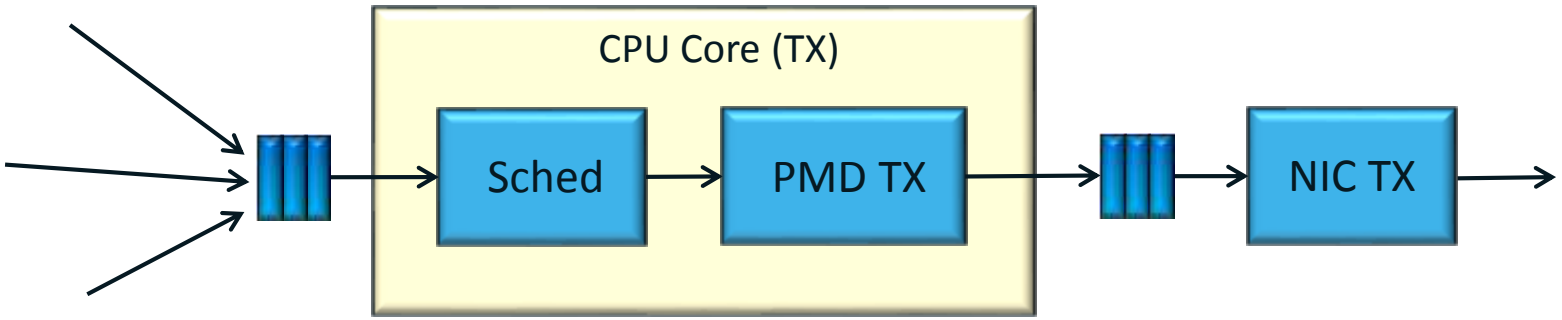


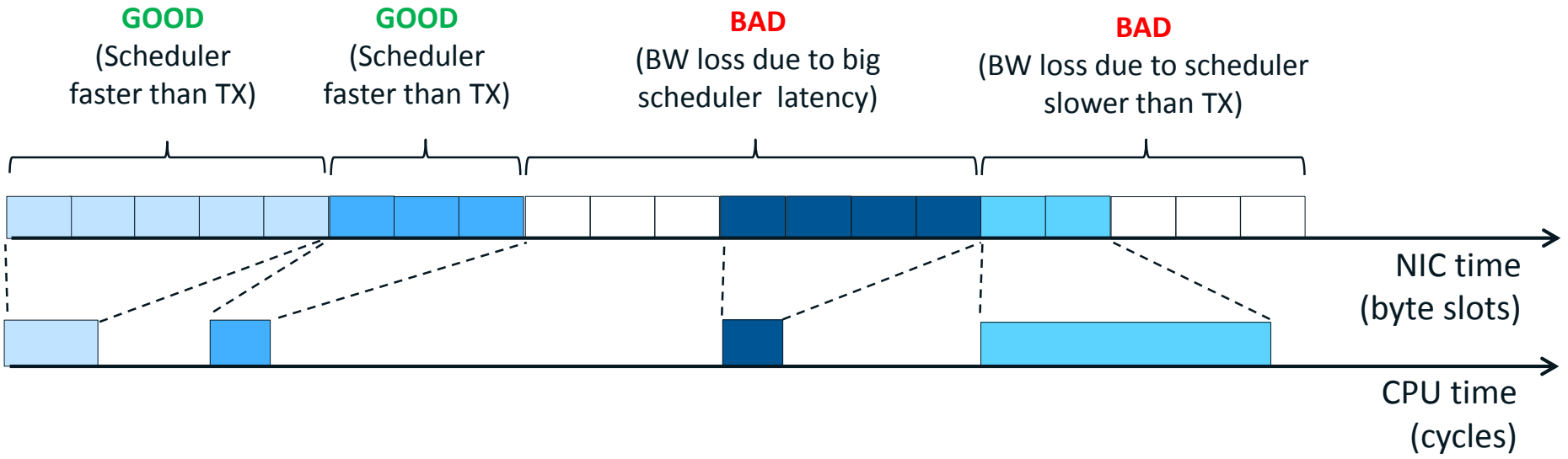
- Each queue has its own bit in the bitmap
  - Bit set by enqueue every time a packet is written to the queue
  - Bit cleared by dequeue every time the queue becomes empty
- Each pipe is represented as a group of 16 bits in the bitmap

Dequeue operation



Timing and Sync





Backup

QoS Keywords Helper